

Veri Madenciliğinde Regresyon Yöntemleri ile Doğalgaz Sektöründe Talep – Tüketim Analizi

Servet Özmen¹, Baha Şen²

¹ Karabük Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Ana Bilim Dalı, Karabük

² Yıldırım Beyazıt Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisliği Bölümü, Ankara
servetozmen@gmail.com, bsen@ybu.edu.tr

Özet: Büyük ölçekli verilerle doğru yöntemler kullanarak gelecekle ilgili tahmin yapmamızı sağlayan kuralları, bilgisayar programlarıyla adlandırma işini veri madenciliği olarak konumlandırabiliriz. Veri madenciliği teknikleri olarak sınıflandırma, bağımlı değişkenler arasındaki ilişkilerin önemini ortaya koyma (regresyon), zaman serileri analizi, özetleme, kümeleme teknikler vardır. Biz regresyon yöntemiyle bu çalışmamızı yapacağız. Doğalgaz sektöründe de gaz tüketimlerinin gelecek tahminlerini yapmak bir dağıtım firması için büyük önem arz etmektedir. Bu önemli tahminleme işlemlerini geriye dönük verilerin ve projeksiyon verileriyle harmanlayarak belirlediğimiz regresyon yöntemleri ile tüketim tahminlemesini yapacağız. Bu sayede elde ettiğimiz sonuçlar, geleceğe dönük enerji tüketimini, yol gösterici niteliğe sahip olmasını ve doğru kararlarla enerji verimliliğini artmasını sağlayacaktır. Bu tahminleme de Sakarya ilinin doğalgaz tüketim verileri ve diğer etken veriler kullanılacaktır.

Anahtar Sözcükler: Veri Madenciliği, Regresyon, Doğalgaz, Tüketim Tahminleme.

Abstract: Large-scale data, which can make estimates about the future rules of using the right methods, computer programs, data mining as a job, can we position the naming. As data mining techniques, classification, regression, time series analysis, summarizing, there are techniques such as clustering. We will do this work the method of regression. To make estimates of the gas consumption of the natural gas sector in the future is of great importance for a distribution company. This is an important retrospective data and the projection data estimation procedures determined by blending with regression methods will forecasting consumption. In this way, our results for the future energy consumption, make the right decisions guiding qualifications will improve energy efficiency. This is the country of Sakarya forecasting of natural gas consumption data and data used in the other factors.

Keywords: Data Mining, Regression, Natural Gas, Consumption, Forecasting

1. Giriş

Doğalgaz sektöründe 2011 yılında LNG ticaret hacmi %10 artarak büyümeye devam etmiştir. LNG ithalatındaki büyüme özellikle nükleer krizle birlikte Japonya ve gelişmekte olan güney yarımküre pazarlarının yanı sıra İngiltere, Güney Kore ve Tayvan gibi gelişmiş pazarlardaki artışlardan kaynaklanmıştır. Avrupa’da LNG talebi yılın ilk yarısında artmış, ancak ikinci yarısında düşüş trendine girmiştir. Asya pazarlarındaki talep artışı ise piyasalardaki fazla LNG miktarını büyük ölçüde azaltmıştır[1].

Küreselleşen dünya da enerji için birçok uygulamalar tahminlemeler yapılması kaçınılmaz durumlara gelmiştir. Bu tahminlemeler ile büyük ülkeler veya şirketler bu süreçte elde edilen sonuçlara göre adım atmaktadırlar.

Türkiye’deki enerji sektörlerinden biri olan doğalgaz da dışa bağımlılığın %95 lerin üstündedir. Bu sebeple gaz dağıtım şirketlerinin günlük gaz çekişlerini kısa dönemlik tahminlemelerle öngörmesi kaçınılmazdır. Bu tahminlemelerde önemli olan geçmiş verilerin bir tabloda elimizde olmasıdır. Tüketim tahminlemesinde birçok faktör sonucu etkilemektedir. Bu veriler ile doğalgaz dağıtım şirketleri öngörülerini belirterek

enerji verimliliği konusunda ülkeye ve dış sermayelere büyük katkılar sağlamaktadır.

Veri madenciliği ile bu elde edilmiş geçmişe dönük verileri harmanlamak çok önemli bir süreç olmuştur.

Bu çalışmada, konutlarda kullanılan doğal gazın ısıtma tüketim dönemlerine ait aylardaki tüketim miktarlarının tahmin edilmesi için geliştirilen regresyon yöntemiyle tanıtılmaktadır. Doğalgaz tüketimiyle, zaman ve derece günlerle ifade edilen hava değişkenleri arasındaki ilişkiler araştırılmakta, ayrıca doğal gaz fiyatı, tüketici fiyat endekslerini kapsayan çeşitli ekonomik göstergelerin doğal gaz kullanımına olan etkisi analiz edilmektedir.

Modeller, Sakarya iline ait gözlem verileri kullanılarak oluşturulmuştur. Elde edilen sonuçlar, zaman ve hava değişkenlerinin yanında tüketicilere yönelik ekonomik göstergelerinde konutlardaki doğal gaz talebi üzerinde belirleyici bir etken olduğunu gözlemlemekteyiz

2. Veri Madenciliği

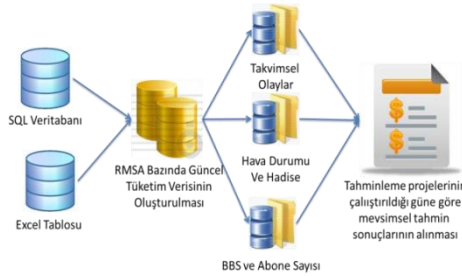
Günümüzde teknolojinin gelişip ucuzlamasından dolayı, her çeşit verinin kaydedilmesiyle enformasyon yığınları, üç boyutlu resimlerde olduğu gibi ilk bakışta hiçbir şeyin anlaşılmadığı karmaşık bilgi çöplüklerine dönüşmüştür. Bu da, verilerin anlamlı bir şekilde

saklanıp gerektiğinde kullanılmasını hem bir ihtiyaç hâline getirmiş, hem de giderek zorlaştırmıştır. Bu noktada, üç boyutlu resimlerin görülebilmesi veya okunabilmesinde bazı metotlar olduğu gibi, veri yığınlarının veri setlerinin verimli bir şekilde okunması ve anlaşılıp kullanılabilmesine yardımcı olan veri madenciliği (Data Mining) veya daha geniş bir yaklaşımla veri yönetimi teknikleri çözüm bulunması konusunda yardımcı olmaktadır.

Veri madenciliği, geleceğe yönelik tahminlerde kullanılacak saklı bilgilerin, birçok veriyi içinde barındıran geniş veri tabanlarından değişik tekniklerle elde edilmesi sanatıdır[2]. Veri madenciliği teknikleri, veri tabanında ve veri ambarlarında (data warehouses) tutulan, bugüne ve geçmiş dönemlere ait verilerden sadece ihtiyacımız olan bilgileri seçip çıkarmaya yarayan yeni bir yaklaşımdır. Bu tekniklerle gelecekte oluşabilecek davranış ve durumlar konusunda tahmin yürütülmekte, beklenen süreçler ortaya çıkmadan kararlar alınarak süreçler yönlendirilebilmektedir.

Bilginin keşfine giden yol olarak da ifade edilen veri madenciliğinde, değişik kaynaklardan veri toplanır, ön işleme tabi tutularak hatalı veriler ayıklanır, eksik-kayıp veriler tamamlanır, veriler ortak bir formata dönüştürülür, gerekli neticelerin elde edilmesi için uygun işlem basamakları (algoritmalar) uygulanır ve neticeler anlaşılır bir şekilde (grafik vb) sunulur. Bu işlemlerin ortak hedefi, eldeki verilerin incelenerek gerçeğe en yakın modele oturtulmasını sağlamaktır. Bu modeller, tahmin edici veya tanımlayıcı hususiyette olabilir.

Veri madenciliği teknikleri olarak sınıflandırma, bağımlı değişkenler arasındaki ilişkilerin önemini ortaya koyma (regresyon), zaman serileri analizi, özetleme, kümeleme, bağlantı kurma gibi teknikler vardır.



Şekil 1. Tahminleme Süreci

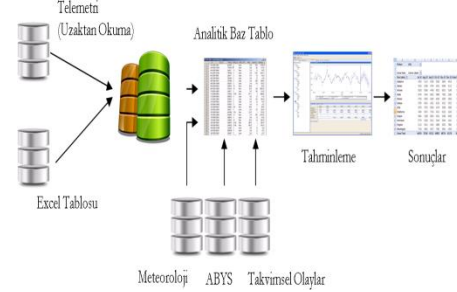
3. Uygulama

3.1 Verilerin Toplanması

Doğal gaz tüketim tahminlemede modeller oluşturulmadan önce, elde edilecek veriler, tahminlerin oluşturulmasında birincil öneme sahiptir. Verilerin doğru ve anlaşılır olması oluşturulacak modelleri ve tahmin sonuçlarını etkileyecektir. Buna göre doğal gaz

tüketimi üzerine kısa dönemde etkiyebilecek değişkenler üzerinde durulmuştur. Bunlar;

- Meteorolojik Veriler
- Abone Bilgileri
- Geçmiş dönem tüketim verileri
- Anlık tüketim verileri
- Takvimsel olaylar şeklindedir.



Şekil 2. Veri Akış Süreci [3]

Modelleme oluşturulurken tahminleme süreci Şekil 2 de görüldüğü gibi olmuştur. Excel tablosu ve SQL veri tabanı üzerinden gelen veriler ile güncel tüketim verileri oluşturulmuş, buna dışarıdan meteoroloji, takvimsel olaylar ve abone bilgileri eklenmiştir

3.2 Meteorolojik Veriler

Gerçekleşen meteoroloji veriler saatlik olarak DMİ 25' den alınmaktadır. Burada iki durum söz konusudur. İlk durum 05 Temmuz 2011 tarihine kadar hazır alınmış DMİ verisi veri tabanına doğrudan aktarılmıştır.

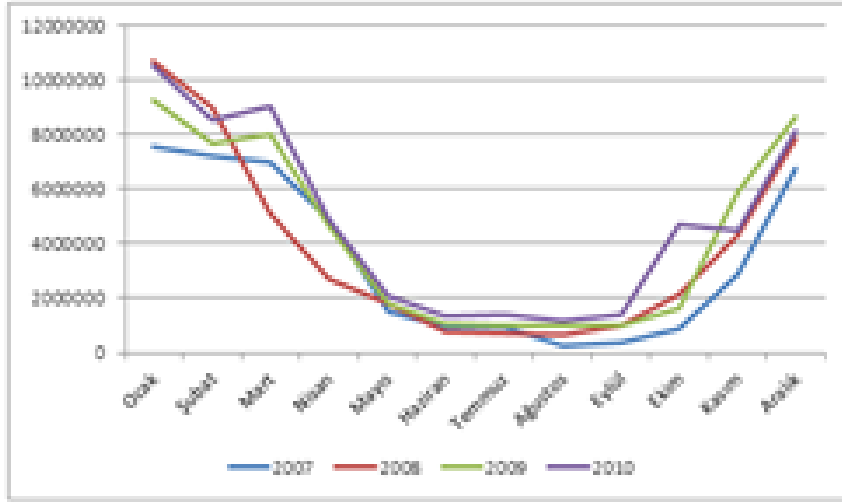
Bu veriler sadece "Adapazarı" istasyonu üzerinden gelmektedir. İşlenmiş olan bu veride en düşük sıcaklık, en yüksek sıcaklık, ortalama sıcaklık, nem, rüzgâr hızı ve rüzgâr yönü bulunmaktadır.

Bu tarihten sonraki (06 Temmuz 2011) meteoroloji verisi ise saatlik ham şekilde DMİ' den sağlanmaktadır. Bu verilerde "Karasu", "Geyve" ve "Adapazarı" istasyonlarından alınan saatlik ortalama sıcaklık, nem, ıslak termometre sıcaklığı, rüzgâr hızı, rüzgâr yönü, hadise bilgisi, basınç ve görüş mesafesi bilgileri yer almaktadır.

Anlık olarak alınan verilerde bulunan mesafesi, hadise bilgisi her zaman alınmamakta, havanın durumuna göre değişmektedir. (Tablo 1)

Tüketimlerin mevsimsel değerler ile rms (reducing metering station – basınç düşürme istasyonu) bazında değişim göstergesi Şekil 3'teki gibidir.

²⁵ Devlet Meteoroloji İşleri



Şekil 3. Meteorolojik Verilerin Tüketimle Orantısı

Tablo 1 - Gerçekleşen Meteoroloji Verisi Örneği

HavaTarih	İlce	T_Ort	T_Min	T_Maks	Nem	Islak Termometre Sıcaklık	Ruzgar Hız	Ruzgar Yonu	Durum	Basınç	Görüş Mesafesi
2011-07-05 01:00:00.000	Karasu	18,3	NULL	NULL	88	18,3	1	220	-99	1012,7	-99
2011-07-05 00:00:00.000	Adapazarı	17,2	NULL	NULL	93	16,2	1	230	HAVA AÇIK	1013,4	15
2011-07-05 00:00:00.000	Geyve	14,7	NULL	NULL	90	13,1	2	70	-99	1015,4	-99
2011-07-05 00:00:00.000	Karasu	18,9	NULL	NULL	90	18,9	3	190	-99	1012,7	-99
2011-07-04 23:00:00.000	Adapazarı	16,8	16,8	17	94	NULL	0,7	199	NULL	NULL	NULL
2011-07-04 22:00:00.000	Adapazarı	NULL	17,6	17,7	NULL	NULL	0,6	150	NULL	NULL	NULL
2011-07-04 21:00:00.000	Adapazarı	17,8	17,8	17,9	92	NULL	0,6	117	NULL	NULL	NULL
2011-07-04 20:00:00.000	Adapazarı	18,8	18,8	19	90	NULL	1	234	NULL	NULL	NULL

3.2 Takvimsel Olaylar

Doğal gaz tüketimi etkileyen diğer bir nokta da takvimsel olaylardır. Takvimsel olaylar resmi tatiller, hafta sonu tatilleri, okulların kapanması, ramazan ve kurban bayramları gibi örnekleri çoğaltılabilecek takvime ve zamana bağlı olarak değişen durumlardır. Burada göz önünde bulundurulacak durumlar; abone türlerine göre;

- İş günü durumu
- Tatil nedeni
- Kampanya durumu

şeklinde oluşturulmuştur. Burada takvimsel olaylar 2006 yılında başlatılmış, 2022 yılına kadar hazırlanmıştır. Takvimsel olaylarda resmi bayramlar girilmiştir. Dini bayramlar belirli tarihe kadar bulunmaktadır. (Tablo 2)

Tatil durumları ise; hafta sonu, resmi tatil, ramazan bayramı, kurban bayramı, seçim & referandum tatili şeklinde oluşturulmuştur.

4. Talep - Tüketim Yöntem Modelleme Analizi

4.1 Regresyon Analizi

Bir ölçüt değişkeni ile bir veya daha fazla sayıda tahmin değişkenleri arasındaki ilgiyi sayısal hale dönüştürmede kullanılan istatistiksel analizdir. Regresyon analizi esas olarak değişkenler arasında ilişkinin niteliğini saptamayı amaçlar.

Tahmin değişkeni olarak bir değişken kullanılırsa basit regresyon, tahmin değişkenleri olarak iki veya daha fazla değişken kullanılırsa çoklu regresyon analizinde söz etmek mümkündür. Amaç her tahmin değişkeninin ölçüt değişkenindeki toplam değişmeye olan katkısının saptanması ve dolayısıyla tahmin değişkenlerinin doğrusal kombinasyonunun değerinden hareketle ölçüt değerinin tahmin edilmesidir. [5].

$$Y_i = (b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n) + e_i$$

IsGunleriID	IsGunleriTarihi	AboneTuru	IsGunuMu	TatilNedeni	Kampanya	KampanyaDerecesi
16537	30.08.2011 00:00	1	1	NULL	NULL	NULL
16538	30.08.2011 00:00	2	1	NULL	NULL	NULL
16539	30.08.2011 00:00	8	0	3	NULL	NULL
16540	30.08.2011 00:00	9	0	3	NULL	NULL
16541	30.08.2011 00:00	10	1	NULL	NULL	NULL
16542	30.08.2011 00:00	5	0	3	NULL	NULL

Tablo 2 – Takvimsel Olaylar Örnek Verisi

Y bağımlı değişken, b₀, regresyon eğrisinin y eksenini kesim noktası, b₁ ilk tahmin değişkeninin X₁ katsayısı, b₂ ikinci tahmin değişkeninin X₂ katsayısı, b₀, regresyon eğrisinin y eksenini kesim noktası, b₁ ilk tahmin değişkeninin X₁ katsayısı, b₂ ikinci tahmin değişkeninin X₂ katsayısı, ... e_i ise i'inci denek için Y'nin tahmin edilen değeriyle gözlenen değeri arasındaki farktır. (Şekil4)

Tarih	Gün	Max	Min	Ort	Nem	Hadise	Ruzgar_H	Ruzgar_Yor	Adapazari_Gerceklesti	Akyazi_Gerceklesti	Camli_Gerceklesti	Hendek_Gerceklesti	Adapazari_Tahm	Akyazi_Tahm
01JAN2012	Pazar	9	8	8	90	KUVYETLURUZGAR	2		239.305	14.022	426.504	48.963	295.337	16.721
02JAN2012	Pazartesi	8	4	7	96	PARÇALI BULLUTLU	1		283.779	16.495	470.288	67.726	32.021	14.096
03JAN2012	Salı	13	2	7	90	SISLI	2	210	280.330	16.888	421.536	63.716	263.539	9.759
04JAN2012	Çarşamba	14	2	8	79	PARÇALI BULLUTLU	2		251.382	14.767	445.304	62.291	289.099	16.572
05JAN2012	Perşembe	12	2	8	91	PARÇALI BULLUTLU	1	220	285.254	16.965	446.832	63.421	281.470	15.342
06JAN2012	Cuma	15	7	11	70	HAYIA KAPALI	1	220	225.887	14.250	379.736	51.054	232.940	12.339
07JAN2012	Cumartesi	7	5	15	96	YAGMURLU	2	330	197.739	13.044	345.544	40.444	262.547	16.982
08JAN2012	Pazar	8	6	8	94	YAGMURLU	2	300	266.872	14.797	397.808	46.398	225.791	14.420
09JAN2012	Pazartesi	6	3	5	96	YAGMURLU	2		279.724	16.397	440.824	70.339	279.902	16.907
10JAN2012	Salı	7	3	5	96	YAGMURLU	2		283.397	19.105	495.696	72.421	293.944	16.289
11JAN2012	Çarşamba	5	3	5	96	KARLA KARISIK YAGMURLU	1	330	288.790	19.511	462.728	71.997	272.123	16.276
12JAN2012	Perşembe	6	3	6	92	HAYIA KAPALI	3	340	295.957	16.620	476.762	72.968	290.895	19.088
13JAN2012	Cuma	8	1	5	95	PARÇALI BULLUTLU	2	230	279.716	19.959	453.368	66.696	295.959	19.950
14JAN2012	Cumartesi	6	2	4	96	YAGMURLU	2	210	260.094	16.864	463.896	62.729	275.431	16.772
15JAN2012	Pazar	5	-1	2	92	HAYIA KAPALI	2		271.694	19.468	506.224	58.708	264.717	16.623
16JAN2012	Pazartesi	4	-1	2	83	HAYIA KAPALI	2		333.020	24.002	606.088	73.853	312.722	20.746
17JAN2012	Salı	4	-2	0	79	KAR SAGANAGI	1	0	343.952	23.712	581.272	76.309	361.869	24.087
18JAN2012	Çarşamba	4	-2	0	76	KAR YAGSIZI	2	220	360.078	24.955	586.216	77.475	339.512	24.177
19JAN2012	Perşembe	5	-2	1	74	HAYIA KAPALI	2	220	361.479	24.295	594.736	79.029	353.314	23.898
20JAN2012	Cuma	9	-1	2	0	HAYIA KAPALI	2	210	360.571	24.696	588.176	73.898	348.278	24.348
21JAN2012	Cumartesi	12	4	8	82	HAYIA KAPALI	19	161	297.795	16.638	500.688	61.457	316.425	24.495
22JAN2012	Pazar	9	5	7	93	YAGMURLU	13	238	246.420	16.704	446.528	54.343	330.322	24.811
23JAN2012	Pazartesi	14	2	10	62	AZ BULLUTLU	11	197	177.033	17.247	411.120	64.955	294.098	19.056
24JAN2012	Salı	11	4	9	87	SAGANAK YAGMURLU	12	172	272.251	16.969	417.362	63.662	295.452	22.202
25JAN2012	Çarşamba	14	7	10	82	HAYIA KAPALI	9	183	296.931	18.444	391.792	63.442	296.939	21.853
26JAN2012	Perşembe	12	7	9	96	KUVYETLURUZGAR	10	179	297.847	17.967	412.200	65.990	289.441	19.906
27JAN2012	Cuma	6	3	2	96	KUVYETLURUZGAR	22	293	323.096	24.107	517.000	76.321	291.270	16.983
28JAN2012	Cumartesi	3	-1	3	96	HAYIA KAPALI	7	33	252.562	16.573	497.640	66.109	297.145	14.698
29JAN2012	Pazar	3	-2	3	93	KUVYETLURUZGAR	10	26	261.798	19.450	520.949	61.019	302.010	23.430
30JAN2012	Pazartesi	1	0	0	92	KUVYETLURUZGAR	16	21	346.648	24.923	604.456	88.676	330.230	24.795
31JAN2012	Salı	0	-3	-2	87	KAR SAGANAGI	18	328	368.911	26.047	628.881	93.997	335.577	25.028
01FEB2012	Çarşamba	1	-3	0	93	KAR SAGANAGI	6	275	368.600	26.353	631.591	92.893	365.803	25.935
02FEB2012	Perşembe	3	-2	-3	88	HAYIA KAPALI	6	36	379.174	26.123	697.464	93.109	379.184	26.108
03FEB2012	Cuma	7	0	1	90	PUSLU	2	120	327.813	23.895	606.496	85.307	359.493	24.054
04FEB2012	Cumartesi	13	3	12	88	PARÇALI BULLUTLU	21	201	230.090	17.92	463.904	62.610	264.427	17.424

Şekil 4. Regresyon Analizi Sonucu Oluşan Veriler

5. Sonuç

Tahminleme sürecinin sonunda sonuçlar oluşmaktadır. Bu sonuçlar günlük tahmin sonuçları olmaktadır. Tahmin sonuçları, yarın ve sonrasını tahmin etmektedir.

Regresyon analiziyle talep – tüketim analizleri hata oranı düşük olarak tahminlemeleri çok yüksek derecede tutmaktadır.

Bu oluşan verileri değişik veri analiz programlarıyla regresyon analizi modellemesini seçerek elde edilebilmektedir.

Geleceğe dönük net raporlar veri kalitesine göre değişiklik gösterebilmektedir.

Bu verilerle raporlar her gün çalışıp modelleme üzerinde yeni verileri işleyebilecektir.

Raporun çalıştığı gündeki tahminler ise kendisinden bir gün önce yapılmış olan tahminler olarak gelmektedir. Böylece geçmişteki günlerde tahminler ortaya rahatlıkla konulabilecek, geliştirme takip edilebilecektir. Tahminler 12 günlük alınmaktadır. Burada önemli olan nokta tahminlerin hava durumunun ilişkisidir.

Üç rapor türü bulunmaktadır. Bunlar;

- Hava durumu, toplam tüketim, toplam tahmin ve hatasını gösteren genel rapor.
- Evsel ve serbest tüketimlerin alt ve üst limitli şekilde rms – a bazlı olduğu detay rapor.
- Hava durumu, gerçekleşen tüketim, tahmini tüketimin evsel ve serbest olarak rms – a bazlı olduğu evsel serbest rapor.

olarak alınmaktadır.

6. Kaynaklar

- [1] Ifpen, " Cedigaz Anticipates New Dynamics In The Global Gas Industry",
<http://www.naturalgaseurope.com>, (2012).
- [2] Vural M., " Veri madenciliği, geleceğe yönelik tahminlerde kullanılacak saklı bilgilerin, birçok veriyi içinde barındıran geniş veri tabanlarından değişik tekniklerle elde edilmesi sanatıdır ", Yüksek Lisans Tezi, 18:269275 (2011).
- [3] Akpınar M. "AGDAŞ, Kısa Dönem Talep Tahminlemesi", Şirket İçi Dokümantasyon , Sakarya (2012).
- [4] Field, Andy "Discovering Statistics, Sage Publications." (2000).
- [5] Kızılaslan, R.,. Forecasting of Short Term and Midterm İstanbul Natural Gas Consumption values by neural network algorithms. Yüksek Lisans Tezi, Fatih Üniversitesi, İstanbul (2008).